Journal of
Applied Crystallography
ISSN 0021-8898

Received 8 April 2004
Accepted 11 August 2004

© 2004 International Union of Crystallography
Printed in Great Britain – all rights reserved

Automatic structure determination from powder data with *EXPO2004*

Angela Altomare,[a] Rocco Caliandro,[a] Mercedes Camalli,[b] Corrado Cuocci,[c] Carmelo Giacovazzo,[a,c]* Anna Grazia Giuseppina Moliterni[a] and Rosanna Rizzi[a]

[a]IC, Sede di Bari, Via G. Amendola 122/o, 70126 Bari, Italy, [b]IC, Sezione di Roma, via Salaria Km. 29.300 Montelibretti Stazione – Roma, Italy, and [c]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

EXPO2004 is the updated version of the *EXPO* program [Altomare *et al.* (1999). *J. Appl. Cryst.* **32**, 339–340]. The traditional steps of the *ab initio* powder solution process are performed automatically: indexing, space-group determination, decomposition of the pattern for extracting the observed structure-factor moduli, structure solution by direct methods, model refinement by Rietveld technique. Special strategies may be applied to improve both the estimates of the extracted structure-factor moduli and the quality of the structure model. In addition, the use of special procedures exploiting available supplementary information on molecular geometry can be successfully adopted. The graphical interface has also been improved.

1. Introduction

Ab initio crystal structure determination from powder data is not a trivial task. Peak overlap, background and preferred orientation can make the crystal structure solution difficult. However, in the past 12 years, owing to the great progress in experimental and computing procedures, the number of structures determined by powder techniques has increased considerably. Many packages have been developed for different purposes: indexation, structure-factor amplitude extraction, structure determination, Rietveld refinement. *EXPO2004* is a program that is applicable over the full pathway from indexation to structure model refinement.

Two methods for structure determination can be followed.

(*a*) The traditional two-stage method (Altomare, Carrozzini *et al.*, 1996; Lasocha & Schenk, 1997) based on the extraction of the structure-factor amplitudes from the experimental pattern and the application of direct or Patterson methods.

(*b*) The global optimization method (Kariuki *et al.*, 1997; David *et al.*, 1998; Engel *et al.*, 1999; Putz *et al.*, 1999; Le Bail, 2001; Favre-Nicolin & Cerny, 2002; Markvardsen *et al.*, 2002; Brodski *et al.*, 2003), which assumes that the structure connectivity is known and selects the best configuration (in terms of bond distances, bond angles and torsion angles) according to the minimum value of a cost function related to the observed pattern.

EXPO2004 uses the two-stage method: full-pattern decomposition is performed according to *EXTRA* (Altomare *et al.*, 1995) and direct methods are applied for solving the structure, according to *SIRPOW92* (Altomare *et al.*, 1994). It preserves the *EXPO* functionality and has been enriched by several new recently developed procedures, as follow.

(i) An automatic peak search routine for locating diffraction peaks.

(ii) Extinction-group determination by statistical analysis of the extracted intensities.

(iii) A random approach in the pattern decomposition for improving the extracted intensities.

(iv) Optimization of the structure model provided by direct methods *via* electron density modification techniques.

(v) In the case of polyhedral coordination, completion of a partial model: missed anions and cations may be located.

(vi) For structures containing heavy atoms, labelling of the Fourier peaks by exploiting the prior information about the atomic connectivity.

(vii) Refinement of the structure model by Rietveld method (Rietveld, 1969).

(viii) A powerful graphical user interface.

The new tools available in *EXPO2004* make the solution process more straightforward, particularly for users who are not so expert in crystallography. The high level of automation, the availability of different strategies and the more user-friendly graphical interface facilitate the successful use of the program with different types of structures.

The main tools of *EXPO2004* and some applications are described below.

2. Indexing procedure

The traditional list of indexing packages [*i.e. ITO* (Visser, 1969), *TREOR90* (Werner *et al.*, 1985), *DICVOL91* (Boultif & Louër, 1991)], has recently been enriched by new programs (Kariuki *et al.*, 1999; Shirley, 2002; Coelho, 2003). With the aim of improving the indexing process, the package *N-TREOR* (Altomare *et al.*, 2000*b*) has been implemented in *EXPO2004*. Its main features are as follow.

(*a*) An efficient peak search procedure, performed by combining two methods, namely the peak profile fit (Huang & Parrish, 1975) and the derivative approach (Savitzky & Golay, 1964; Sonneveld & Visser, 1975; Huang, 1988).

(*b*) A more intensive search for the correct cell parameters.

(*c*) An automatic least-squares *d* refinement of the most plausible cell.

(*d*) A user-friendly graphical interface.

In order to increase the probability of success, the unit-cell search has been made more automatic and effective, as follows.

(i) The default values of *TREOR90* for the maximum investigated volume and axis of the cell (2000 Å^3 and 25 Å, respectively) are increased in *N-TREOR* to 4000 Å^3 and 35 Å when the maximum observed *d* value is larger than 10 Å.

(ii) When no plausible cell is found at the end of a default run, *N-TREOR* performs a new search with larger tolerance and increased values for the maxima of the Miller indices.

(iii) The reduced and the conventional cells are automatically calculated *via* the program *REDUCT* (Westdahl & Werner, 1995).

(iv) At the end of the indexing run, more plausible solutions can be suggested. The solution having the maximum M'_{20} figure of merit is indicated as the most probable one, where $M'_{20} = (7 - N_p)M_{20}$; N_p is the number of cell parameters to be determined and M_{20} is the de Wolff figure of merit (de Wolff, 1968). The selected conventional cell is automatically refined and characterized.

(v) After a first run of *N-TREOR*, positive and negative origin shifts are alternatively applied in order to undo the effects of a possible zero shift in the 2θ Bragg angles. The most probable zero shift is selected as that corresponding to the cell with the largest value of M'_{20}.

For a comparison between *TREOR90* and *N-TREOR* performances, the reader is referred to the work of Altomare *et al.* (2000*b*).

3. Space-group determination

The unequivocal determination of the crystalline space group using powder X-ray diffraction data is not straightforward: usually more than one space group may be compatible with the experimental peak distribution. An efficient probabilistic approach has been proposed recently by Markvardsen *et al.* (2001). In *EXPO2004*, we have introduced a new probabilistic approach (Altomare *et al.*, 2004) which is able to provide, on a quantitative basis, a probability value for each extinction symbol compatible with the previously established lattice symmetry. The user can choose the extinction symbol *via* the graphical interface, which also presents a list of the feasible space groups. Furthermore, the space group may be selected graphically and used immediately in the subsequent calculations.

An example of graphical output is presented in Fig. 1 for UTM1 (Plévert *et al.*, 1999). When the highest figure of merit for the extinction group is accepted, the compatible space groups are supplied on the right-hand side of the screen. However, the user can choose any space group *via* a text box.

4. The full-pattern decomposition procedure

The full-pattern decomposition is performed *via* the Le Bail *et al.* (1988) algorithm, which has the great advantage that the intensity values can be optimized if additional prior information is available (Altomare, Carrozzini *et al.*, 1996; Altomare, Foadi *et al.*, 1996; Altomare, Giacovazzo *et al.*, 1999; Altomare *et al.*, 1998; Carrozzini *et al.*, 1997; Rodriguez-Carvajal, 1990; Estermann *et al.*, 1992). Its main limit is the tendency to equipartition the intensity of strongly overlapped reflections. A new extraction procedure (Altomare *et al.*, 2001) has been introduced in *EXPO2004*, aiming at breaking the equipartition tendency. For each cluster of strongly overlapping reflections and large overall intensity, the amplitudes of the individual reflections are randomly varied under the condition that the overall intensity remains constant and are submitted to the Le Bail algorithm. The residual profile value R_p (between the observed and the calculated profile of each cluster) is calculated and the decompositions characterized by the lowest R_p values are selected as the most reliable. The amplitudes so obtained, together with those not belonging to the selected clusters, are submitted to the direct phasing procedure. This strategy increases the computing time, but the probability of obtaining trials with low average phase error is higher.

5. Structure model optimization

Ab initio crystal structure solution of proteins with data at atomic resolution became a viable process when reciprocal-space methods (*e.g.* tangent procedures or equivalent techniques) were coupled with real-space methods (*e.g.* electron modification techniques) [see Rappleye *et al.* (2002) for the program *SnB*, and Sheldrick *et al.* (2001) for the program *SHELX-D*]. Owing to the low value of the ratio 'number of independent observations/number of parameters to refine', the standard Fourier recycling procedure in *EXPO* provides a rough and often incomplete structure model. To overcome this drawback, a real-space-based approach, originally applied to the *ab initio* crystal structure solution of proteins (Giacovazzo & Siliqi, 1997; Burla *et al.*, 2000), has been introduced in *EXPO2004* (Altomare, Cuocci *et al.*, 2002). It consists of the EDM and RFR steps.

The EDM (electron density modification) step consists of six cycles $\rho \rightarrow \{\varphi\} \rightarrow \rho$: each half-cycle $\rho \rightarrow \{\varphi\}$ is carried out by inverting a quite small percentage of the map, about 2.5%, corresponding to the pixels with largest electron density values. In the second half-cycle, a variable number (say N_i) of reflections is used, which increases with the cycle order.

The RFR (real-space Fourier recycling) step follows the EDM approach and consists of ten cycles $\{\varphi\} \rightarrow \rho \rightarrow \{\varphi\}$. For the first four cycles, the structure-factor calculation is performed assigning the lightest non-hydrogen atomic species to the located electron density peaks. For the remaining six cycles, the peak labelling is carried out, taking into account the peak intensity, the chemical cell content and all the atomic species of the crystal structure.

The gain (in terms of phase error) provided by the above procedure is often non-negligible: for example, for the UTM1 structure, the tangent process ends with an error of 44°, which is reduced to 38° after the application of EDM and RFR. The corresponding results for GAPO (Meden *et al.*, 1997) are 28° and 19°, respectively.

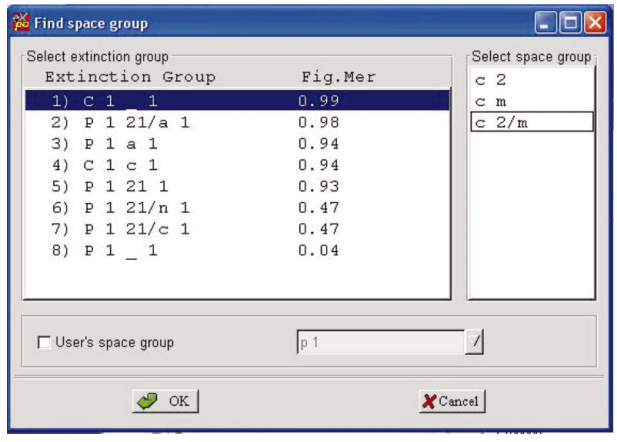


Figure 1
The result of the space-group determination step for UTM1.

6. The use of connectivity information in the peak labelling routine

At the end of a Fourier procedure, one or more molecular fragments are obtained: the atomic species are usually assigned according to the intensities of the electron density peaks. In this situation, spurious but intense peaks, incorrectly labelled, can hinder the recovery of the correct structural model. A new algorithm was implemented in *EXPO2004* (Altomare, Giacovazzo *et al.*, 2002), aiming at using the crystal chemical information that is usually available *a priori* to improve peak labelling. The prior information is constituted by: (i) the heavy atomic species present in the unit cell; (ii) the number of heavy atoms in the asymmetric unit; (iii) the expected heavy atom to heavy atom distances and the expected number of heavy atoms surrounding each heavy atom at an expected distance; (iv) the expected heavy atom to light atom distances. The efficiency of the algorithm decreases with the number of heavy atomic species; in *EXPO2004* it may be applied to crystal structures containing only one or two different heavy atomic species (with $Z > 10$).

7. The use of polyhedral coordination information

When a crystal structure contains both heavy and light atoms, the E map usually locates the heavy atoms; the recovery of the light atoms may be very difficult even with sophisticated EDM or least-squares procedures. To recover the full structure, the approach described by Altomare *et al.* (2000*a*) and by Giacovazzo *et al.* (2002) has been implemented in *EXPO2004* (see also Grosse-Kunstleve *et al.*, 1997): it is able to combine the information on the cation coordination, usually available *a priori*, with Monte Carlo methods. The completion process generates anion positions in a random way, but all compatible with the prior information on the coordination polyhedra (tetrahedral or octahedral) and on the expected cation–anion distances. The procedure is able to complete the structure, both when all the cations have been positioned and when some of them are missed. The configuration corresponding to the best agreement between the experimental and the calculated pattern is chosen as the most reliable.

8. The model refinement by Rietveld technique

The main features of Rietveld refinement (Rietveld, 1969; Young, 1993) in *EXPO2004* are as described below.

(*a*) Parameters defining unit cell, background, profile function, preferred orientation, 2θ-zero correction, scale factor, atomic positions, thermal factor and atomic site occupancy can be optimized by using one of the following techniques: (i) Gauss–Newton algorithm combined with a line search procedure (Dennis & Schnabel, 1996) (the default choice); (ii) Levenberg–Marquardt algorithm (Marquardt, 1963); (iii) standard Gauss–Newton algorithm. The user can graphically set the damping factors to be applied to the shifts.

(*b*) The available peak-shape functions are: Pearson VII, pseudo-Voigt, Lorentzian and Gaussian (Wiles & Young, 1981).

(*c*) Background can be calculated by choosing one of the following types of functions: Young's polynomial, Chebyshev polynomial and cosine Fourier series (Wiles & Young, 1981).

(*d*) Starting values of the profile parameters are provided by *EXTRA*. Agreement factors on profile and structure-factor moduli are calculated to check the refinement process: χ^2, R_p, R_{pw}, R_F, R_{Bragg}. Background-subtracted R_p and R_{pw} values are also determined.

(*e*) The observed and calculated profiles, and their difference, are plotted cycle by cycle of the refinement.

(*f*) A preliminary Rietveld refinement can be carried out in an automatic way. It consists of two steps: firstly, lattice, scale factor, background function, 2θ-zero correction, profile and asymmetry adjustable parameters are refined; secondly, the atomic parameters and the isotropic thermal factors are optimized.

(*g*) The number of observations can be increased in order to make the refinement process more stable by using information on restraints on distances and angles (Waser, 1963; Baerlöcher, 1982; Blanc & Paciorek, 2001). The graphical interface enables the selection, by mouse click, of the atoms involved in the restraints.

In Fig. 2, we present an example concerning the Rietveld application to the structure AND2 (Chernyshev *et al.*, 1999): the observed, calculated and difference profiles are plotted, in addition to the variation of R_{pw} cycle by cycle.

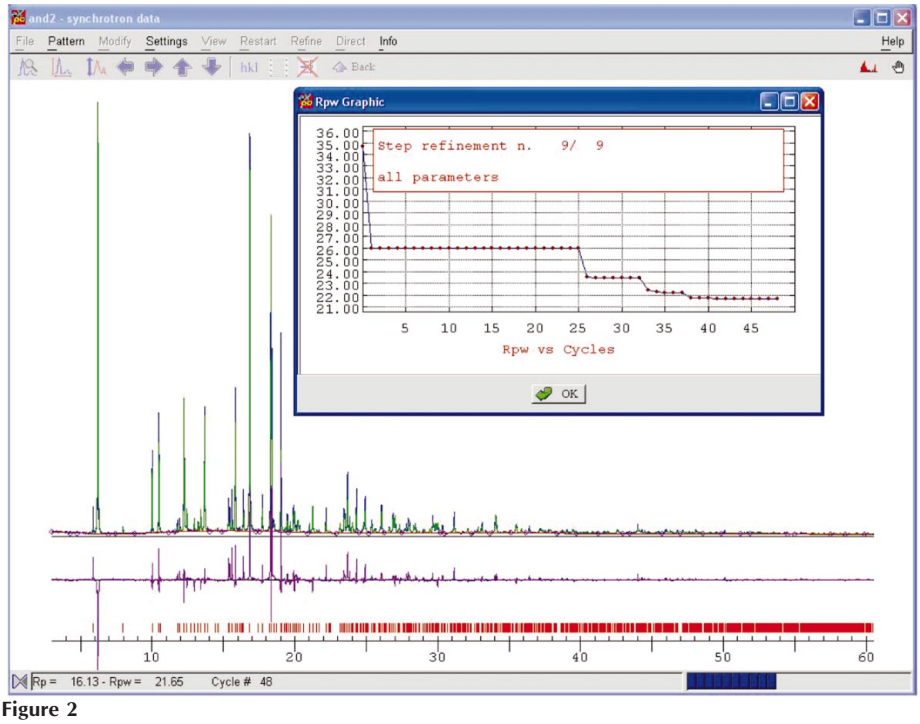


Figure 2
Rietveld refinement of AND2: the progress of the refinement is depicted cycle by cycle.

9. Conclusions

The *EXPO* program has been strengthened by new computational and graphical features. The result of such work, the program *EXPO2004*, is able to perform all the steps traditionally used for the crystal structure solution: indexation of the experimental pattern, space-group determination, decomposition of the pattern in single intensities, phasing *via* direct methods, refinement of the structural model *via* Rietveld techniques. A high level of automation has been secured (the program evaluates the quantity of information available at each critical point and makes its own choices), but the user can

apply special strategies when the solution is not attained by the default run.

10. Program features

10.1. Hardware and software environment

The program runs on a PC or workstation, under Windows 9*x*/2000/Me/NT/XP, Linux, Unix. On a PC, some DDLs are necessary and they are supplied in the distribution kit. On Unix platforms, a Fortran 90 and C compiler are needed, together with the GTK/GDK/GLIB libraries.

10.2. Program specification

There are no limits on the number of counts in the pattern. The maximum number of reflections is 3000. The maximum number of atoms in the asymmetric unit is 500. The maximum number of different atomic species is 8.

10.3. Program availability and documentation

EXPO2004 is available *via* http://www.ic.cnr.it. The software is free of charge for non-profit organizations. After the Licence Agreement has been signed and faxed to the Institute of Crystallography, the program can be downloaded. The installation instructions and the user manual are accessible *via* the Web. Documentation about the program itself and on the usage of the graphical interface is provided in HTML format.

The authors are indebted to Gianluca Cascarano and Mariarosaria Mallamo for their help in the development of the graphical environment.

References

Altomare, A., Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Rizzi, R. (1999). *J. Appl. Cryst.* **32**, 339–340.
Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A. & Polidori, P. (1994). *J. Appl. Cryst.* **27**, 435–436.
Altomare, A., Burla, M. C., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Polidori, P. (1995). *J. Appl. Cryst.* **28**, 842–846.
Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *J. Appl. Cryst.* **37**, 957–966.
Altomare, A., Carrozzini, B., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R. (1996). *J. Appl. Cryst.* **29**, 667–673.
Altomare, A., Cuocci, C., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R. (2002). *J. Appl. Cryst.* **35**, 182–184.
Altomare, A., Foadi, J., Giacovazzo, C., Guagliardi, A. & Moliterni, A. G. G. (1996). *J. Appl. Cryst.* **29**, 674–681.
Altomare, A., Foadi, J., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Burla, M. C. & Polidori, G. (1998). *J. Appl. Cryst.* **31**, 74–77.
Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R. (1999). *J. Appl. Cryst.* **32**, 963–967.
Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R. (2000*a*). *J. Appl. Cryst.* **33**, 1305–1310.
Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000*b*). *J. Appl. Cryst.* **33**, 1180–1186.
Altomare, A., Giacovazzo, C., Ianigro, M., Moliterni, A. G. G. & Rizzi, R. (2002). *J. Appl. Cryst.* **35**, 21–27.
Altomare, A., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2001). *J. Appl. Cryst.* **34**, 704–709.
Baerlöcher, Ch. (1982). *The X-ray Rietveld System*. Zurich: Institut für Kristallographie und Petrographie.
Blanc, E. & Paciorek, W. (2001). *J. Appl. Cryst.* **34**, 480–483.
Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987–993.
Brodski, V., Peschar, R. & Schenk, H. (2003). *J. Appl. Cryst.* **36**, 239–243.
Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2000). *Acta Cryst.* A**56**, 451–457.
Carrozzini, B., Giacovazzo, C., Guagliardi, A., Rizzi, R., Burla, M. C. & Polidori, G. (1997). *J. Appl. Cryst.* **30**, 92–97.
Chernyshev, V. V., Fitch, A. N., Sonneveld, E. J., Kurbakov, A. I., Makarov, V. A. & Tafenko, V. A. (1999). *Acta Cryst.* B**55**, 554–562.
Coelho, A. A. (2003). *J. Appl. Cryst.* **36**, 86–95.
David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.* **8**, 931–932.
Dennis, J. E. & Schnabel, R. B. Jr (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia: SIAM.
Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.
Estermann, M.A., McCusker, L. B. & Baerlocher, C. (1992). *J. Appl. Cryst.* **25**, 539–543.
Favre-Nicolin, V. & Cerny, R. (2002). *J. Appl. Cryst.* **35**, 734–743.
Giacovazzo, C., Altomare, A., Cuocci, C., Moliterni, A. G. G. & Rizzi, R. (2002). *J. Appl. Cryst.* **35**, 422–429.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, C. (1997). *J. Appl. Cryst.* **30**, 985–995.
Huang, T. C. (1988). *Aust. J. Phys.* **41**, 201–212.
Huang, T. C. & Parrish, W. (1975). *Appl. Phys. Lett.* **27**, 123–124.
Kariuki, B. M., Serrano-Gonzales, H., Johnston, R. L. & Harris, K. D. M. (1997). *Chem. Phys. Lett.* **280**, 189–195.
Kariuki, B. M., Belmonte, S. A., McMahon, M. I., Johnston, R. L., Harris, K. D. M. & Nelmes, R. J. (1999). *J. Synchrotron Rad.* **6**, 87–92.
Lasocha, W. & Schenk, H. (1997). *J. Appl. Cryst.* **30**, 909–914.
Le Bail, A. (2001). *Mater. Sci. Forum*, **378–381**, 65–70.
Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.
Markvardsen, A. J., David, W. I. F., Johnson, J. C. & Shankland, K. (2001). *Acta Cryst.* A**57**, 47–54.
Markvardsen, A. J., David, W. I. F. & Shankland, K. (2002). *Acta Cryst.* A**58**, 316–326.
Marquardt, D. (1963). *SIAM J. Appl. Math.* **11**, 431–441.
Meden, A., Grosse-Kunstleve, R. W., Baerlocher, C. & McCusker, L. B. (1997). *Z. Kristallogr.* **212**, 801–807.
Plévert, J., Yamamoto, K., Chiari, G. & Tatsumi, T. (1999) *J. Phys. Chem. B*, **103**, 8647–8649.
Putz, H., Schön, J. C. & Jansen, M. (1999). *J. Appl. Cryst.* **32**, 864–870.
Rappleye, J., Innus, M., Weeks, C. M. & Miller, R. (2002). *J. Appl. Cryst.* **35**, 374–376.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Rodriguez-Carvajal, J. (1990). *Abstracts of the Satellite Meeting on Powder Diffraction of the XV Congress of the IUCr*, p. 127, Toulouse, France.
Savitzky, A. & Golay, M. J. E. (1964). *Anal. Chem.* **36**, 1627–1639.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2001). *International Tables for Crystallography*, Vol. F, pp. 333–345.
Shirley, R. (2002). *Acta Cryst.* A**58** (Suppl.), C16.
Sonneveld, E. J. & Visser, J. W. (1975). *J. Appl. Cryst.* **8**, 1–7.
Visser, J. W. (1969). *J. Appl. Cryst.* **2**, 89–95.
Young, R. A. (1993). *The Rietveld Method*. Oxford University Press.
Waser, J. (1963). *Acta Cryst.* **16**, 1091–1094.
Werner, P.-E., Eriksson, L. & Westdahl, M. (1985). *J. Appl. Cryst.* **18**, 367–370.
Westdahl, M. & Werner, P.-E. (1995). *World Directory of Powder Diffraction Programs*, Release 3, edited by S. Gorter & D. Smith. Chester: IUCr.
Wiles, D. B. & Young, R. A. (1981). *J. Appl. Cryst.* **14**, 149–151.
Wolff, P. M. de (1968). *J. Appl. Cryst.* **1**, 108–113.